I claim:

1.   A method for reducing word sense ambiguities in a sentence, based on thematic

prediction, said method comprising the steps of:

5   a.   receiving an input sentence consisting of a sequence of part-of-speech tagged

                words;

        b.   creating a sequence of sense tagged words from said received sequence of part-of-

                speech tagged words, each of said senses further being theme tagged;

        c.   predicting a set of one or more probable themes associated with said created

10   sequence of sense-tagged words;

        d.   weighting each of said one or more probable themes from said predicted set, and

        e.   reducing sense ambiguities by eliminating remotely probable senses or selecting

                highly probably senses from said weighted set of one or more probable themes.

2.   A method for reducing word sense ambiguities in a sentence, based on thematic

15   prediction, as in claim 1, wherein said set of predicted one or more probable themes for

        said input sentence belongs to a predefined set of coarse grain themes.

3.   A method for reducing word sense ambiguities in a sentence, based on thematic

        prediction, as in claim 1, wherein said step of predicting said set of one or more probable

        themes comprises the following steps:

20   a.   searching a database and identifying any pre-stored words in said input sentence;

b.   assigning a theme for each of said identified pre-stored words in said input sentence;

c.   accessing a lexicon and identifying one or more themes associated with words in said input sentence, and

d.   outputting all of said assigned and identified themes for said input sentence.

4.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 3, wherein said lexicon comprises a limited set of words for a given language, and each of said words are associated with one or more parts-of-speech, and each of said parts-of-speech is associated with one or more senses, and each of said one or more senses is assigned one or more themes out of a set of pre-defined themes.

5.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 3, wherein said database is accessible over a network.

6.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 8, wherein said network is any of the following: wide area network (WAN), local area network (LAN), Internet, or wireless networks.

7.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 1, wherein said step of weighting each of said predicted set of one or more probable themes further comprises calculating a theme score, said theme score depending on:

a.   a coefficient whose value depends on parts-of-speech associated with each word in said input sentence, and

b.   number of different words with a similar theme in said input sentence.

- 19 -

8.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 1, wherein said step of reducing sense ambiguities is eliminated when more than one of said predicted set of probable themes have the same weighting and said weighting is the highest one among the set of predicted themes.

9.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 1, wherein said step of reducing sense ambiguities is performed only if the number of words in said input sentence possessing a dominant theme is at least equal to ¼ the total number of words in said input sentence.

10.   A method for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 1, wherein said reduced sense ambiguities are used as inputs to a natural language processing system.

11.   A system for reducing word sense ambiguities in a sentence, based on thematic prediction, said system comprising:

a thematic predictor receiving an input sentence comprising a sequence of part-of-speech tagged words and outputting a sequence of sense tagged words and a set of one or more predicted themes associated with said sequence of tagged words;

a thematic scorer weighting each of said set of one or more predicted themes, and

a thematic word sense disambiguator reducing sense ambiguities by eliminating remotely probable senses or selecting highly probable senses from said weighted set of one or more probable themes.

12.   A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said thematic predictor further searches a database and

identifies any pre-stored words in said input sentence and assigns a theme for each of said identified pre-stored words in said input sentence.

13. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 12, wherein said pre-stored words and themes in said database are updated regularly.

14. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 12, wherein said database is accessible over a network.

15. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 14, wherein said network is any of the following: wide area network (WAN), local area network (LAN), Internet, or wireless networks.

16. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said thematic predictor further accesses a lexicon and identifies one or more themes associated with words in said input sentence.

17. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said lexicon comprises a limited set of words for a given language, and each of said words are associated with one or more parts-of-speech, and each of said parts-of-speech is associated with one or more senses, and each of said one or more senses is assigned one or more themes out of a set of pre-defined themes.

18. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said system further comprises a morphological analyzer for stemming each word in said input sentence and annotating each of said

stemmed words with at least one part of speech tag to form said sequence of part-of-speech tagged words.

19. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said system further comprises an interface for displaying said weighted one or more predicted themes and said eliminated sense ambiguities as a result of disambiguation.

20. A system for reducing word sense ambiguities in a sentence, based on thematic prediction, as in claim 11, wherein said thematic scorer further scores each of said predicted set of one or more probable themes by calculating a theme score, said theme score depending on:

   a. a coefficient whose value depends on parts-of-speech associated with each word in said input sentence, and

   b. number of different words with a similar theme in said input sentence.

21. An article of manufacture comprising a computer user medium having computer readable code embodied therein which reduces word sense ambiguities in a sentence, based on thematic prediction, said medium comprising:

   computer readable program code receiving an input sentence consisting of a sequence of part-of-speech tagged words;

   computer readable program code creating a sequence of sense tagged words from said received sequence of part-of-speech words, each of said senses further being theme tagged;

computer readable program code predicting a set of one or more probable themes associated with said created sequence of sense-tagged words;

computer readable program code weighting each of said predicted set of one or more probable themes, and

computer readable program code reducing sense ambiguities by eliminating remotely probable senses or selecting highly probably senses based on said weighted set of one or more probable themes.

22.   An article of manufacture comprising a computer user medium having computer readable code embodied therein which reduces word sense ambiguities in a sentence, based on thematic prediction, as in claim 21, wherein computer readable code predicting said set of one or more probable themes further comprises:

a.   computer readable code searching a database and identifying any pre-stored words in said input sentence;

b.   computer readable code assigning a theme for each of said identified pre-stored words in said input sentence;

c.   computer readable code accessing a lexicon and identifying one or more themes associated with words in said input sentence, and

d.   computer readable code outputting all of said assigned and identified themes for said input sentence.

23.   An article of manufacture comprising a computer user medium having computer readable code embodied therein which reduces word sense ambiguities in a sentence, based on thematic prediction, as in claim 21, wherein said computer readable code further provides

- 23 -

for an interface for displaying said weighted one or more predicted themes and said

eliminated sense ambiguities as a result of disambiguation.